

Enumeration of Structured Flowcharts

EDWARD A. BENDER

University of California-San Diego, La Jolla, California

AND

JON T. BUTLER

Northwestern University, Evanston, Illinois

Abstract. An analysis of structured flowcharts is presented, where size is measured by the number, n , of decision nodes (IF-THEN-ELSE and DO-WHILE nodes). For all classes of structured flowcharts considered, the number of charts is approximately, $cn^{-3/2}\gamma^n$, for large n , where c and γ are parameters that depend on the class. It is also shown that most large flowcharts consist of a short sequence of basic charts (IF-THEN-ELSE and DO-WHILE charts). The average length of such sequences is 2.5.

Categories and Subject Descriptors: D.2.2. [Software Engineering]: Tools and techniques—*structured programming; top-down programming*; G.2.1 [Discrete Mathematics]: Combinatorics—*generating functions; permutations and combinations; recurrences and difference equations*

General Terms: Performance, Theory

Additional Key Words and Phrases: BJ-charts, D-charts, decision node, flowchart enumeration, sequency, structured programming

1. Introduction

During the past 20 years, progress has been made toward putting the theory of programming on a firm mathematical basis. McCabe [6] proposed that program complexity be measured by the cyclomatic number, which is one plus the number of decision nodes in the program flowchart. Studies have been done to determine how well such measures perform in practice [1] and to what extent they predict the number of programming errors [7].

In a much cited paper, Dijkstra [4] proposed that the GO TO statement be eliminated in high-level languages, since it allowed unnecessarily complicated program structures. It was suggested, rather, that transfer from one part of the program to another be done only through the use of a small set of constructs. One such set consists of the DO-WHILE, IF-THEN-ELSE, and SEQUENCE. Flow-

The research of the first author was supported by the National Science Foundation under Grant MCS 79-27060. The research of the second author was supported by a National Research Council Senior Postdoctoral Associateship while the second author was tenured at the Naval Postgraduate School, Monterey, California.

Authors addresses: E. A. Bender, Department of Mathematics, University of California-San Diego, La Jolla, CA 92093; J. T. Butler, Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60201

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

© 1985 ACM 0004-5411/85/0700-0537 \$00.75

Report Documentation Page			Form Approved OMB No. 0704-0188		
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE JUL 1985		2. REPORT TYPE		3. DATES COVERED	
4. TITLE AND SUBTITLE Enumeration of Structured Flowcharts				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Northwestern University, Department of Electrical Engineering and Computer Science, Evanston, IL, 60201				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT An analysis of structured flowcharts is presented, where size is measured by the number, n, of decision nodes (IF-THEN-ELSE and DO-WHILE nodes). For all classes of structured flowcharts considered, the number of charts is approximately, $cn-?/*y?$, for large n, where c and y are parameters that depend on the class. It is also shown that most large flowcharts consist of a short sequence of basic charts (IF-THEN-ELSE and DO-WHILE charts). The average length of such sequences is 2.5.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 12	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

charts formed from these are called *D-charts*, after Dijkstra. Böhm and Jacopini [3] showed that any algorithm can be implemented as a D-chart.

Knuth [5] and others have advocated less restrictive constructs to avoid additional computation sometimes required by the constraints of D-charts. One less restrictive construct is the DO-WHILE with one or more midloop exits. Together with the IF-THEN-ELSE and SEQUENCE, these form the basis of the BJ_m -charts, after Böhm and Jacopini. BJ_m -charts form a hierarchy. That is, BJ_m -charts include BJ_{m-1} -charts, which, in turn, include BJ_{m-2} -charts, and so on. At the lower end are the BJ_1 -charts, which are identical to D-charts.

Although there is a better understanding of program structure and complexity, the full range of the programming process has not been explored. We believe that further progress depends on a better understanding of the universe from which all programs are produced. That is, when a programmer designs a program, a choice is made from the set of all programs. Thus, there is the question of whether a programmer produces programs that correspond to a random choice or whether there is a bias. Our results indicate the latter is true.

We begin by enumerating BJ_m -charts. It is shown that the number of n -node charts is approximately $cn^{-3/2}\gamma^n$, where c and γ are constants near 0.9 and 9, respectively. We show that, in large D-charts, approximately 56 percent of the decision nodes are IF-THEN-ELSE, while the remaining 44 percent are DO-WHILE. In large charts, where one or more midloop exits are allowed in DO loops, the converse is true; a larger percentage of nodes is DO-WHILE. We also investigate the composition of large flowcharts. That is, any flowchart corresponds to a sequence of basic charts, which are the IF-THEN-ELSE and DO-WHILE charts. It is shown that large flowcharts, on the average, consist of a small number (two or three) of basic charts. Long sequences (of more than eight basic charts) are very rare.

2. Preliminaries

A *flowchart* is a directed graph where nodes represent code segments and arcs represent transfer of control. A *decision node* is a node with indegree 1 and outdegree 2, representing a code segment from which control proceeds to one of two points, depending on the value of a predicate. Figure 1a shows a decision node and Figure 1b shows the representation used in this paper.

Note that interchanging T (true) and F (false) in Figure 1a is identical to replacing P with its complement. Since all possible predicates are allowed, it is appropriate to just omit T and F, as is done in Figure 1b. The other type of node, the *reconvergent node*, has indegree 2 and outdegree 1, as shown in Figure 1c. It represents a code segment to which flow of control converges. We deal in this paper exclusively with one-input one-output flowcharts. It follows, therefore, that the number of decision nodes is identical to the number of reconvergent nodes. However, this paper focuses on the decision node and, from now on, the term *node* will refer to a decision node.

2.1. D-CHARTS. D-charts are built up from smaller D-charts in one of three ways, as shown in Figure 2. Figure 2a shows the IF-THEN-ELSE construct. A chart of the form shown in Figure 2a is called an *IF-THEN-ELSE flowchart*, and η is called an *IF-THEN-ELSE node*. Interchanging G_1 and G_2 leaves the chart unchanged, since it is the same as complementing the predicate at η . A flowchart of the form in Figure 2b is called a *DO-WHILE flowchart* and η is called a *DO-*

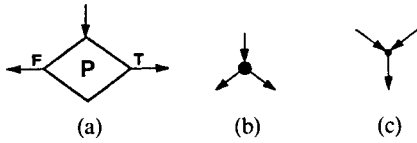
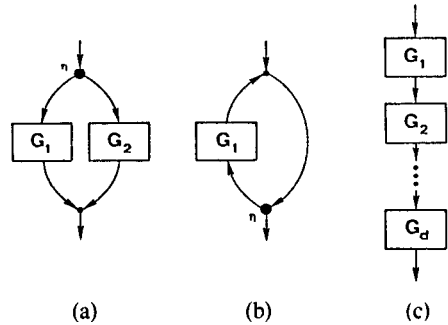


FIG. 1. Flowchart nodes. (a) and (b) Decision node. (c) Reconvergent node.

FIG. 2. The IF-THEN-ELSE, DO-WHILE, and SEQUENCE constructs. (a) IF-THEN-ELSE. (b) DO-WHILE. (c) SEQUENCE.



WHILE node. Figure 2c shows the *SEQUENCE* flowchart, built up by concatenating IF-THEN-ELSE or DO-WHILE flowcharts.

Definition 1. D-Chart:

- (1) A directed arc is a D-chart.
- (2) For $h \geq 1$, the SEQUENCE of G_1, G_2, \dots, G_h is a D-chart where G_i is an IF-THEN-ELSE or DO-WHILE chart with subcharts that are D-charts for $1 \leq i \leq h$.

Let the *sequency* of a D-chart be the value of h . It is convenient to let the sequency of a directed arc be 0. Charts of sequency 1 are precisely the IF-THEN-ELSE and DO-WHILE charts. All charts with 0, 1, and 2 decision nodes are shown in Figure 3. Figure 3a shows a chart with sequency 0, Figure 3b shows charts with sequency 2, Figure 3c single-node charts with sequency 1 and Figure 3d double-node charts with sequency 1.

Not all flowcharts are D-charts. Figure 4 is not realizable as any combination of IF-THEN-ELSE, DO-WHILE, and SEQUENCE for any integer $j > 1$.

2.2. BJ_m -CHARTS. The construct shown in Figure 4 was introduced by Böhm and Jacopini [3] as a model for the midloop exit of a DO loop. A flowchart of the form shown is called a *DO-WHILE-WITH- j -EXITS* flowchart. $\eta_1, \eta_2, \dots, \eta_j$ are called *DO-WHILE-WITH- j -EXITS* nodes. Note that the DO-WHILE and DO-WHILE-WITH-1-EXIT flowcharts are identical.

Definition 2. BJ_m -Chart (m a positive integer or ∞):

- (1) A directed arc is a BJ_m -chart.
- (2) For $h > 1$, the SEQUENCE of G_1, G_2, \dots, G_h is a BJ_m -chart, where each G_i is either an IF-THEN-ELSE chart with BJ_m -subcharts or a DO-WHILE-WITH- j -EXITS chart ($j \leq m$) with BJ_m -subcharts.

The *sequency* of a BJ_m -chart is 0 (respectively, h) if it satisfies Part 1 (respectively, 2) of Definition 2.

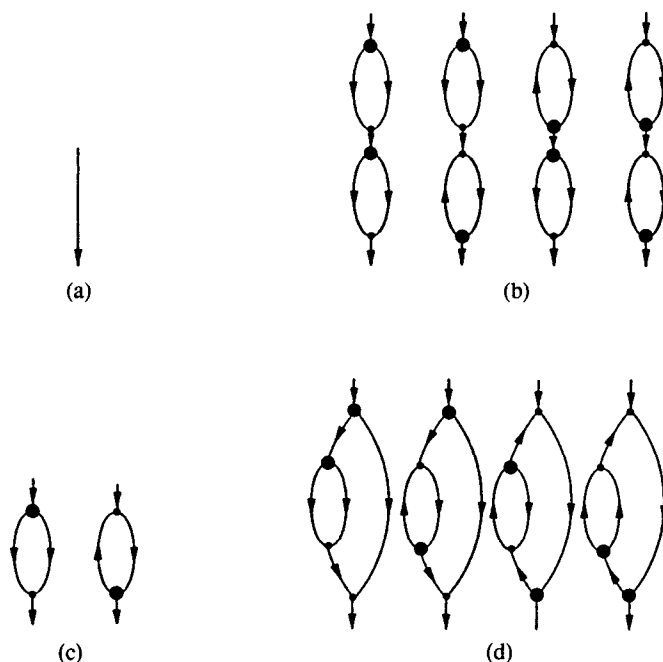
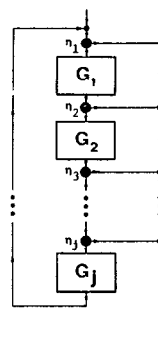


FIG. 3. All D-charts with 0, 1, and 2 decision nodes. (a) Flowchart with sequency 0. (b) Flowcharts with sequency 2. (c) Single-node flowcharts with sequency 1. (d) Double-node flowcharts with sequency 1.

FIG. 4. The DO-WHILE-WITH- j -EXITS construct.



3. Flowchart Enumeration by Generating Functions

Let $f_m(n)$ be the number of BJ_m -charts with n decision nodes. For example, from Figure 3 we have for $BJ_1 = \text{D-charts}$, $f_1(0) = 1$, $f_1(1) = 2$, and $f_1(2) = 8$. Let $F_m(x)$ be the generating function for BJ_m -charts; that is,

$$F_m(x) = f_m(0) + f_m(1)x + f_m(2)x^2 + \dots + f_m(i)x^i + \dots, \quad (1)$$

where x is a formal variable. An expression for $F_m(x)$ will be obtained from $ITE_m(x)$ and $DWW_m(x)$, the generating functions for the number of BJ_m -charts which are IF-THEN-ELSE and DO-WHILE-WITH- m -EXITS flowcharts, respectively.

Consider $DWW_m(x)$ first. Distinct choices for G_1, G_2, \dots, G_j in Figure 4 result in distinct DO-WHILE-WITH- j -EXITS flowcharts. Thus,

$$DWW_m(x) = \sum_{j=1}^m x^j F_m^j(x). \quad (2)$$

Here, $x^j F_m^j(x)$ corresponds to the contribution of the DO-WHILE-WITH- j -EXITS flowcharts. $F_m(x)$ is the contribution from each of the subcharts, while x^j is the contribution from the j decision nodes. From (2),

$$\text{DWW}_m(x) = \frac{x F_m(x)(1 - x^m F_m^m(x))}{1 - x F_m(x)}, \quad (3)$$

where x^∞ is interpreted as 0.

Consider $\text{ITE}_m(x)$. If distinct choices for subcharts G_1 and G_2 resulted in a distinct IF-THEN-ELSE flowchart, then $\text{ITE}_m(x)$ would be $x F_m^2(x)$. The generating function for those charts with $G_1 = G_2$ is $x F_m(x^2)$. Since interchanging G_1 and G_2 produces the same flowchart,

$$\text{ITE}_m(x) = \frac{x F_m^2(x) - x F_m(x^2)}{2} + x F_m(x^2) = \frac{x F_m^2(x) + x F_m(x^2)}{2} \quad (4)$$

Since a BJ_m -chart is a SEQUENCE of some number i of IF-THEN-ELSE and DO-WHILE-WITH- m -EXITS flowcharts, the generating function for BJ_m -charts can be written as

$$F_m(x) = \sum_{i=0}^{\infty} [\text{ITE}_m(x) + \text{DWW}_m(x)]^i. \quad (5)$$

Thus,

$$F_m(x) = [1 - \text{ITE}_m(x) - \text{DWW}_m(x)]^{-1}. \quad (6)$$

Substituting (3) and (4) into (6) yields

$$F_m(x) = \left[1 - \frac{x F_m^2(x)}{2} - \frac{x F_m(x^2)}{2} - x F_m(x) \frac{1 - x^m F_m^m(x)}{1 - x F_m(x)} \right]^{-1} \quad (7)$$

This function equation can be used to obtain the coefficients of $F_m(x)$ recursively, as follows. Let $P_d(x)$ be a polynomial of degree d that agrees with $F_m(x)$ through terms of degree d . Using it in place of $F_m(x)$ in the right side of (7) and truncating at degree $d+1$ produces $P_{d+1}(x)$.

Table I shows the results. It is interesting to note that, for a fixed n , the number of n node BJ_m -charts that are also BJ_{m-1} is quite substantial for $m > 2$. However, as will be seen shortly, for a large enough n , BJ_{m-1} -charts represent an arbitrarily small fraction of BJ_m -charts.

4. Asymptotic Approximations to the Number of BJ_m -Charts

In this section, we derive an asymptotic approximation for $f_m(n)$. We will use Theorem 5 of Bender [2]:

THEOREM 1 [2, p. 502]. Assume the power series $w(x) = \sum_{n=0}^{\infty} a_n x^n$ with nonnegative coefficients satisfies $G(x, w) \equiv 0$ in which w and x are related implicitly. Suppose there exist real numbers $r > 0$ and $s > a_0$ such that

- (i) for some $\Delta > 0$, $G(x, w)$ is analytic for $|x| < r + \Delta$ and $|w| < s + \Delta$,
- (ii) $G(r, s) = G_w(r, s) = 0$,
- (iii) $G_x(r, s) \neq 0$ and $G_{ww}(r, s) \neq 0$, and
- (iv) if $|x| \leq r$, $|w| \leq s$, and $G(x, w) = G_w(x, w) = 0$, then $x = r$ and $w = s$.

Then,

$$a_n \sim cn^{-3/2} \gamma^n \left(\text{i.e., } \lim_{n \rightarrow \infty} \frac{a_n}{cn^{-3/2} \gamma^n} = 1 \right) \quad (8)$$

TABLE I. THE NUMBER $f_m(n)$ OF BJ_m-CHARTS VERSUS THE NUMBER n OF DECISION NODES

	$f_1(n)$	$f_2(n)$	$f_3(n)$	$f_4(n)$	$f_5(n)$	$f_\infty(n)$
n	D = BJ ₁	BJ ₂	BJ ₃	BJ ₄	BJ ₅	BJ _∞
0	1	1	1	1	1	1
1	2	2	2	2	2	2
2	8	9	9	9	9	9
3	43	53	54	54	54	54
4	258	347	359	360	360	360
5	1,682	2,463	2,584	2,598	2,599	2,599
6	11,529	18,358	19,526	19,680	19,696	19,697
7	82,058	141,959	153,026	154,602	154,793	154,812
8	600,320	1,127,755	1,231,851	1,247,453	1,249,517	1,249,770
9	4,487,352	9,150,633	10,126,949	10,278,871	10,300,132	10,303,076
10	34,120,281	75,508,725	84,658,561	86,124,597	86,338,009	86,370,016

for

$$c = \left(\frac{rG_x(r, s)}{2\pi G_{ww}(r, s)} \right)^{1/2} \quad (9)$$

and

$$\gamma = \frac{1}{r}. \quad (10)$$

With w playing the role of $F_m(x)$ in (7), we have,

$$G(x, w) = x^{m+1}w^{m+2} + \frac{x^2w^4}{2} - \frac{xw^3}{2} + \left(\frac{x^2F_m(x^2)}{2} - 2x \right) w^2 + \left(x + 1 - \frac{xF_m(x^2)}{2} \right) w - 1 = 0. \quad (11)$$

Differentiating G with respect to w yields

$$G_w(x, w) = (m+2)x^{m+1}w^{m+1} + 2x^2w^3 - \frac{3xw^2}{2} + (x^2F_m(x^2) - 4x)w + x + 1 - \frac{xF_m(x^2)}{2}. \quad (12)$$

Given $x = r$ and $w = s$, satisfying the conditions of Theorem 1, we have,

$$G_x(r, s) = (m+1)r^ms^{m+2} + rs^4 - \frac{s^3}{2} + (r^3F'_m(r^2) + rF_m(r^2) - 2)s^2 + \left(1 - r^2F'_m(r^2) - \frac{F_m(r^2)}{2} \right) s \quad (13)$$

and

$$G_{ww}(r, s) = (m+2)(m+1)r^{m+1}s^m + 6r^2s^2 - 3rs + r^2F_m(r^2) - 4r. \quad (14)$$

The values of r , s , $F_m(r^2)$, and $F'_m(r^2)$ were found numerically as in [2, Example 7.2], resulting in the asymptotic expressions for $F_m(n)$ shown in Table II. The error in the approximations for $n = 10$ ranges from 6.1 percent for $f_1(10)$ to 7.4 percent for $f_\infty(10)$.

TABLE II. ASYMPTOTIC APPROXIMATIONS TO $f_m(n)$, THE AVERAGE NUMBER $\epsilon(n)$ OF IF-THEN-ELSE AND DO-WHILE-WITH- m -EXITS NODES, THE FRACTION $\mu(n)$ OF CHARTS THAT ARE IF-THEN-ELSE, DO-WHILE-WITH- m -EXITS, AND HAVE SEQUENCY > 1 , AND THE AVERAGE SEQUENCY $\sigma_m(n)$ OF FLOWCHARTS

Asymptotic approximation for	$BJ_1 = D$	BJ_2	BJ_3	BJ_4	BJ_5	BJ_∞
$f_m(n)$	$0.3888n^{-3/2}8.849^n$	$0.3863n^{-3/2}9.594^n$	$0.3828n^{-3/2}9.716^n$	$0.3814n^{-3/2}9.738^n$	$0.3810n^{-3/2}9.741^n$	$0.3808n^{-3/2}9.742^n$
$\epsilon_{ITE}(n)$	0.5409	0.4613	0.4448	0.4411	0.4403	0.4401
$\epsilon_{DWW}(n)$	0.4591	0.5387	0.5552	0.5589	0.5597	0.5599
$\mu_{ITE}(n)$	0.2015	0.1828	0.1794	0.1787	0.1786	0.1785
$\mu_{DWW}(n)$	0.1130	0.1423	0.1498	0.1516	0.1520	0.1521
$\mu_{SQCY>1}(n)$	0.6855	0.6749	0.6708	0.6697	0.6695	0.6694
$\sigma_m(n)$	2.4933	2.4422	2.4226	2.4172	2.4158	2.4154

5. Properties of Large BJ_m -Charts

In this section, we calculate how many nodes there are of various types, how many flowcharts there are with various sequences, and the average sequency of n -node flowcharts, where $n \rightarrow \infty$. In the following analysis, we consider generating functions $A(x)$ with explicit representations. Thus, a special case of Darboux's theorem (stated in [2, Theorem 4]) will be useful.

THEOREM 2 [2, p. 498]. *Suppose $A(x) = \sum_{n=0}^{\infty} a_n x^n$ is analytic near 0 and can be written in the form*

$$A(x) = h(x) + \left(1 - \frac{x}{\alpha}\right)^{-\omega} g(x), \quad (15)$$

where α is the only singularity of $A(x)$ such that $|\alpha| < |\beta|$ for all other singularities β of $A(x)$. Further, if g and h are analytic near α , $\omega \neq 0, -1, -2, \dots$, and $g(\alpha) \neq 0$, then

$$\alpha_n \sim \frac{g(\alpha)n^{\omega-1}}{\Gamma(\omega)\alpha^n}, \quad (16)$$

where $\Gamma(\omega)$ is the gamma function.

Theorem 2 shows that the asymptotic behavior of coefficients of x^n in $A(x)$ is determined by the singularity of least magnitude.

In deriving Theorem 1, Bender [2, p. 505] shows that $w(x)$ has a power series expansion in powers of $(1 - x/r)^{1/2}$ near $x = r$ beginning with

$$s - \left(\frac{2rG_x(r, s)}{G_{ww}(r, s)}\right)^{1/2} \left(1 - \frac{x}{r}\right)^{1/2}. \quad (17)$$

This applies to $w = F_m(x)$ by the result of the previous section. Near $(x, w) = (r, s)$, $G_w(x, w)$ has a power series expansion beginning with

$$G_w(r, s) + G_{xw}(r, s)(x - r) + G_{ww}(r, s)(w - s). \quad (18)$$

Combining (17), (18), and $G_w(r, s) = 0$, we see that $G_w(x, w)$ has a power series expansion in $(1 - x/r)^{1/2}$ beginning with

$$- (2rG_x(r, s)G_{ww}(r, s))^{1/2} \left(1 - \frac{x}{r}\right)^{1/2}. \quad (19)$$

5.1. AVERAGE NUMBER OF NODES OF VARIOUS TYPES. Consider now the calculation of the average number of nodes of various types. Let $\epsilon_{ITE}(n)$ and $\epsilon_{DWW}(n)$ be the fraction of nodes in n -node charts that are type IF-THEN-ELSE and DO-WHILE-WITH- m -EXITS, respectively. Since there are exactly two types of nodes.

$$\epsilon_{ITE}(n) + \epsilon_{DWW}(n) = 1. \quad (20)$$

We solve first for $\epsilon_{ITE}(n)$ and then use (20) to find $\epsilon_{DWW}(n)$. Let $F_m(x, y) = \sum a_{np} x^n y^p$, where a_{np} is the number of n -node flowcharts p of which are IF-THEN-ELSE nodes. If we differentiate $F_m(x, y)$ with respect to y , and set $y = 1$, we get $P_m(x) = \sum (a_{n1} + 2a_{n2} + \dots + na_{nn})x^n$. The coefficient of x^n divided by the number of BJ_m -charts with n nodes is the average number of nodes that are IF-

THEN-ELSE. $F_m(x, y)$ satisfies the functional equation

$$F_m(x, y) = \left[1 - \frac{xyF_m^2(x, y)}{2} - \frac{xyF_m(x^2, y^2)}{2} - xF_m(x, y) \frac{1 - x^m F_m^m(x, y)}{1 - xF_m(x, y)} \right]^{-1}, \quad (21)$$

which has a form similar to (7) except that x in the factors corresponding to IF-THEN-ELSE nodes is replaced by xy . We have

$$P_m(x) = \left. \frac{\partial F_m(x, y)}{\partial y} \right|_{y=1} = \frac{N(x, w)}{G_w(x, w)}, \quad (22)$$

where

$$N(x, w) = \frac{-x^2 w^4}{2} + \frac{xw^3}{2} - \left(\frac{x^2 F_m(x^2)}{2} + x^2 P_m(x^2) \right) w^2 + \left(\frac{x F_m(x^2)}{2} + x P_m(x^2) \right) w, \quad (23)$$

and $G_w(x, w)$ is given by (12). By (19) and (22), the series expansion of $P_m(x)$ in powers of $(1 - x/r)^{1/2}$ begins with

$$-N(r, s)(2rG_x(r, s)G_{ww}(r, s))^{-1/2} \left(1 - \frac{x}{r} \right)^{-1/2}. \quad (24)$$

By Theorem 2, (9), and (24), the average number of IF-THEN-ELSE nodes, $n_{\text{ITE}}(n)$ satisfies

$$\begin{aligned} n_{\text{ITE}}(n) &\sim - \frac{N(r, s)(2rG_x(r, s)G_{ww}(r, s))^{-1/2} n^{-1/2} \gamma^n}{(rG_x(r, s)/2\pi G_{ww}(r, s))^{1/2} n^{-3/2} \gamma^n \Gamma(1/2)} \\ &\sim - \frac{N(r, s)}{rG_x(r, s)} n. \end{aligned} \quad (25)$$

Thus,

$$\epsilon_{\text{ITE}}(n) \sim - \frac{N(r, s)}{rG_x(r, s)}. \quad (26)$$

Table II shows $\epsilon_{\text{ITE}}(n)$ and $\epsilon_{\text{DWW}}(n) = 1 - \epsilon_{\text{ITE}}(n)$ as $n \rightarrow \infty$ for the various flowcharts. It is interesting to note (by comparing D and BJ_∞) that allowing arbitrary midloop exits only increases the fraction of DO-WHILE-type modes by 10 percent.

5.2. NUMBER OF IF-THEN-ELSE AND DO-WHILE-WITH- m -EXITS CHARTS. We turn now to the calculation of flowchart sequency. Recall that the sequency of a flowchart f is the number of IF-THEN-ELSE and DO-WHILE-WITH- m -EXITS charts concatenated to form f . First consider charts of sequency 1. Such charts are of two types. The generating function of n -node IF-THEN-ELSE charts is given in (4). Using (17) and reasoning as before, we find that $\mu_{\text{ITE}}(n)$, the fraction of charts that are type IF-THEN-ELSE is given as

$$\mu_{\text{ITE}}(n) \sim rs. \quad (27)$$

In a similar fashion, $\mu_{\text{DWW}}(n)$, the fraction of charts which are type DO-WHILE-WITH- m -EXITS can be derived from (3), yielding,

$$\mu_{\text{DWW}}(n) \sim \frac{mr^{m+2}s^{m+1} - (m+1)r^{m+1}s^m + r}{(1 - rs)^2}. \quad (28)$$

Since $\mu_{ITE}(n)$ and $\mu_{DWW}(n)$ account for all n -node flowcharts with sequency 1, the fraction of charts, $\mu_{SQCY>1}(n)$ with sequency 2 or more is just

$$\mu_{SQCY>1}(n) = 1 - \mu_{ITE}(n) - \mu_{DWW}(n). \quad (29)$$

Table II shows $\mu_{ITE}(n)$, $\mu_{DWW}(n)$, and $\mu_{SQCY>1}(n)$ as $n \rightarrow \infty$. Surprisingly, μ_{ITE}/μ_{DWW} is not close to $\epsilon_{ITE}/\epsilon_{DWW}$.

5.3. DISTRIBUTION OF FLOWCHARTS WITH RESPECT TO SEQUENCY. We extend these results to count the flowcharts with various sequencies. Specifically, the number of n -node BJ_m -flowcharts with sequency h is the coefficient of x^n in

$$S(x) = [ITE_m(x) + DWW_m(x)]^h.$$

Let $\delta_{hm}(n)$ be the fraction of n -node BJ_m -charts with sequency h . Reasoning as before,

$$\begin{aligned} \delta_{hm}(n) \sim h & \left[\frac{rs^2}{2} + \frac{rF_m(r^2)}{2} + \frac{rs(1 - r^m s^m)}{1 - rs} \right]^{h-1} \\ & \times \left[rs + \frac{mr^{m+2}s^{m+1} - (m+1)r^{m+1}s^m + r}{(1 - rs)^2} \right]. \end{aligned} \quad (30)$$

Table III lists the values of $\delta_{hm}(n)$ for $n \rightarrow \infty$, $1 \leq h \leq 8$, and $m = 1, 2, 3, 4, 5$, and ∞ .

Figure 5 shows a plot of $\delta_{hm}(n)$ versus h for two flowchart classes, $D = BJ_1$ charts and BJ_∞ -charts. It can be seen from Table III and Figure 5 that the average sequency is between 2 and 3. These averages $\sigma_m(n)$, can be calculated by multiplying (30) by h and summing over h . The results are shown in Table II.

6. Concluding Remarks

We approach the problem of enumerating flowcharts from two points of view. In Section 3, generating functions are used to calculate the exact number of flowcharts with 1 to 10 decision nodes. In Section 4 we develop asymptotic approximations to the number of BJ_m -charts and find that it grows as $cn^{-3/2}\gamma^n$ for large n , where c and γ are constants. The error in the approximation is only 7 percent at $n = 10$.

We observe that BJ_{m-1} -charts represent a vanishing fraction of BJ_m -charts as $n \rightarrow \infty$. This is in contrast to the case of small n ($1 \leq n \leq 10$). For example, for $n = 10$, BJ_2 -charts represents 89 percent of the set of BJ_3 -charts.

An analysis of the types of nodes in large charts shows, for example, that in $D = BJ_1$ -charts, IF-THEN-ELSE nodes represent about 56 percent of the total, while DO-WHILE nodes represent the remaining 44 percent. It is interesting to compare this with the fact that there are almost twice as many IF-THEN-ELSE charts as there are DO-WHILE charts. Thus, in a SEQUENCE of large D -charts, we would expect the node types that determine the subcharts to be distributed in a different proportion than the set of all nodes.

Another characteristic investigated is sequency, the number of basic subcharts in a SEQUENCE that composes the flowchart. The average sequency is about 2.5. Thus, long chains of basic charts (IF-THEN-ELSE and DO-WHILE) are rare in structured flowcharts.

We have observed that programmers tend to produce programs with large sequency. It would seem therefore that either

TABLE III. THE PROPORTION $\delta_{hm}(n)$ OF BJ_m-CHARTS FOR $n \rightarrow \infty$ VERSUS SEQUENCY h

Sequency h	$\delta_{h1}(n)$	$\delta_{h2}(n)$	$\delta_{h3}(n)$	$\delta_{h4}(n)$	$\delta_{h5}(n)$	$\delta_{h\infty}(n)$
	D = BJ ₁	BJ ₂	BJ ₃	BJ ₄	BJ ₅	BJ _{∞}
1	0.315	0.325	0.329	0.330	0.331	0.331
2	0.276	0.279	0.281	0.281	0.281	0.281
3	0.182	0.180	0.179	0.179	0.179	0.179
4	0.107	0.103	0.102	0.102	0.102	0.102
5	0.059	0.055	0.054	0.054	0.054	0.054
6	0.031	0.029	0.028	0.028	0.028	0.028
7	0.016	0.014	0.014	0.014	0.014	0.014
8	0.008	0.007	0.007	0.007	0.007	0.007
Average sequency σ_m	2.4933	2.4422	2.4226	2.4172	2.4158	2.4154

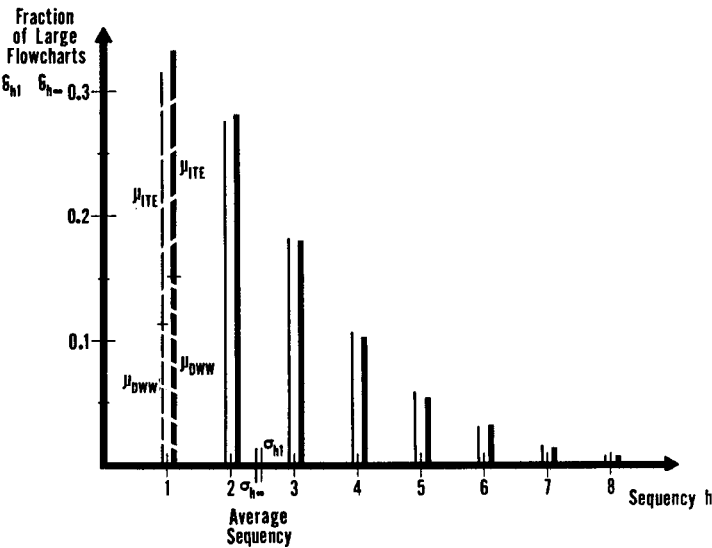


FIG. 5. Asymptotic approximations to $\delta_{h1}(n)$ and $\delta_{h\infty}(n)$, the fraction of n -node flowcharts with sequency h , for D = BJ₁-charts and for BJ _{∞} -charts, respectively: D = BJ₁-charts, —; BJ _{∞} -charts, - - -. Note that μ_{ITE} , μ_{DWW} , δ_{h1} , and $\delta_{h\infty}$ are the values to which $\mu_{ITE}(n)$, $\mu_{DWW}(n)$, $\delta_{h1}(n)$, and $\delta_{h\infty}(n)$ are asymptotically constant.

- (a) the programs we are solving are inherently sequential, or
- (b) IF-THEN-ELSE and DO-WHILE constructs are not enough to help us overcome our sequential organizational tendencies.

A better understanding of these issues may have important consequences for the design of programming languages and our abilities to handle parallel processing.

ACKNOWLEDGMENT. Special thanks are due Paul Ham for conversations that inspired topics in this paper.

REFERENCES

1. BAKER, A. L., AND ZWEBEN, S. H. A comparison of measures of control flow completely. *IEEE Trans. Softw. Eng. SE-6*, 6 (Nov. 1980), 506-511.

2. BENDER, E. A. Asymptotic methods in enumeration. *SIAM Rev.* 16 (1974), 485-515. Errata, 18 (1976), 292.

3. BÖHM C., AND JACOPINI, G. Flow-diagrams, Turing machines, and languages with only two formation rules. *Commun. ACM* 9 (May 1966), 366-371.
4. DIJKSTRA, E. W. Go to statement considered harmful. *Commun. ACM* 11 (Mar. 1968), 147-148.
5. KNUTH, D. E. Structured programming with GOTO statement. *Comput. Surv.* 6, 4 (Dec. 1974), 261-301.
6. MCCABE, T. J. A complexity measure. *IEEE Trans. Softw. Eng.* SE-2, 4 (Dec. 1976), 308-320.
7. SCHNEIDEWINDE, N. J. AND HOFFMAN, H. M. An experiment in software error data collection and analysis. *IEEE Trans. Softw. Eng.* SE-5, 3 (May 1979) 276-286.

RECEIVED JULY 1983; REVISED JANUARY 1985; ACCEPTED FEBRUARY 1985